

© Jan Christoph Meister. All rights reserved. Version 03.09.2003. Contact: [mail@jcmeister.de](mailto:mail@jcmeister.de)

## Projekt Computerphilologie<sup>[1]</sup>

### Über Geschichte, Verfahren und Theorie rechnergestützter Literaturwissenschaft

JAN CHRISTOPH MEISTER

Universität Hamburg

Computerphilologie, was auch immer man sich darunter im Einzelnen vorzustellen hat, setzt offenbar zweierlei voraus: Philologie als eine institutionalisierte Textwissenschaft, wie wir sie seit dem 19. Jahrhundert kennen, und natürlich – Computer. Erinnern wir uns: kommerzielle Großrechner kamen in den frühen 50er Jahren des 20. Jahrhunderts auf den Markt; Literaturwissenschaftler machten ihre erste Bekanntschaft mit Computern allerdings in der Regel nicht vor Anfang der 80er Jahre, als in größeren Bibliotheken die traditionellen Zettelkataloge allmählich durch elektronische Systeme abgelöst wurden. Mitte der 80er begann dann die rasante Erfolgsgeschichte von Apple, Mac und PC. Der Siegeszug der sogenannten ‚Personal Computer‘ hat auch den Literaturwissenschaftlern viele neue Möglichkeiten eröffnet und uns zudem unabhängig gemacht von den universitären Rechenzentren, deren IBM-Mainframes ohnehin weitgehend von Naturwissenschaftlern und Statistikern genutzt wurden. Wenn man ein Stichdatum für den Beginn der Computerphilologie ansetzen wollte, so läge dieses also irgendwo zwischen 1950 und 1985.

Richtig und dennoch falsch: denn das *Projekt* Computerphilologie (nachfolgend: CP) gibt es schon wesentlich länger als den Computer selbst. Bereits 1726 - also vor mehr als 275 Jahren! – z.B. erschien dieser Bericht über ein an der Akademie von Lagado betriebenes Projekt:

The Professor...said, perhaps I might wonder to see him employed in a Project for improving speculative Knowledge by practical mechanical Operations.... Every one knew how laborious the usual Method is of attaining to Arts and Sciences; whereas by his Contrivance, the most ignorant Person...may write Books in Philosophy, Poetry, Politicks, Law, Mathematicks and Theology, without the least Assistance from Genius or Study.

He then led me to the Frame, about the Sides whereof all his Pupils stood in Ranks. It was Twenty Foot square [and] composed of several Bits of Wood, about the Bigness of a Dye....They were all linked together by slender Wires. These Bits of Wood were covered on every Square with Paper pasted on them; and on these Papers were written all the Words of their Language in their several Moods, Tenses, and Declensions, but without any Order....

The Pupils at his Command took each of them hold of an Iron Handle, whereof there were Forty fixed round the Edges of the Frame; and giving them a sudden Turn, the whole Disposition of Words was

entirely changed. He then commanded Six and Thirty of the Lads to read the several Lines softly as they appeared upon the Frame; and where they found three or four Words together that might make Part of a Sentence, they dictated to the four remaining boys....

Six hours a-Day the young Students were employed in this Labour; and the Professor shewed me several Volumes in large Folio already collected, of broken Sentences, which he intended to piece together; and out of those rich Materials to give the World a complete Body of all Arts and Sciences.

Jonathan Swift: *Gulliver's Travels*. (1726) Boston, New York  
1995:173-175

Der Berichterstatter ist eine der berühmtesten Figuren der Weltliteratur, der von dem irischen Satiriker Jonathan Swift auf Erkundungsfahrt geschickte Lemuel Gulliver. Ihm verdanken wir auch eine Lithographie, die eine Teilansicht der geschilderten ‚Literary Engine‘ liefert:

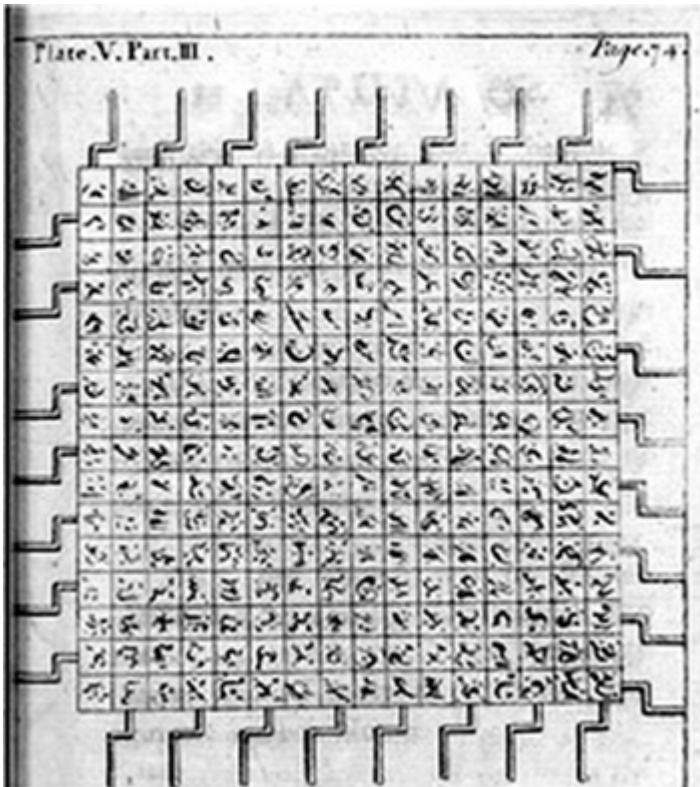


Abbildung 1: Jonathan Swifts ‚Literary Engine‘.  
 Plate V in J.Swift, *Travels into several remote nations of  
 the world. By Lemuel Gullive‘*; Motte-Edition von 1726.  
 Abbildung nach <http://www.jaffebros.com/lee/gulliver/bancroft/10.jpeg> (19.08.2003)

Swifts Satire zielte hier auf zweierlei. Allgemein galt sie dem im 18. Jahrhundert dominanten enzyklopädischen Wissenschaftstypus, der die Idee des aufklärerischen Rationalismus naiv auf das Projekt einer mechanischen Sammlung des Wissens reduzieren wollte. Konkret hingegen attackierte Swift den Philosophen Gottfried Wilhelm Leibniz, der die Möglichkeit einer Mechanisierung deduktiver Prozesse und der automatischen Generierung von sprachlichen Äußerungen behauptet und damit die Tradition der sogenannten ‚Lullischen Kunst‘ aufgegriffen hatte. Diese ‚Ars Magna‘ war von dem 1232 auf Mallorca geborenen Philosophen Raimundus Lullus begründet worden, der eine mechanische Begriffskombinatorik entwarf, mit der man aus etablierten Konzepten auf formalem Wege neue sinnhaltige Begriffe generieren und so zu neuen rationalen Einsichten gelangen sollte. Eines der kombinatorischen ‚Tools‘ von Lullus war der nachfolgend abgebildete Begriffszirkel, der die semantisch-logischen Beziehung benachbarter Konzepte visualisiert:



Abbildung 2: Raimundus Lullus' erster Zirkel aus der *Ars generalis ultima*. Abbildung nach <http://www.bu.edu/wcp/Papers/Cogn/CognWild.htm> (19.08.2003)

Das absurde Projekt der lagadonischen *Literary Engine* wie seine durchaus ernst zu nehmenden Vorbilder bei Lullus, Leibniz, Bruno und anderen machen auf eine Erwartungshaltung aufmerksam, die noch das gegenwärtige Informationszeitalter bestimmt.<sup>[2]</sup> Was dem Zeitgenossen des 18. Jahrhunderts die *Maschine*, ist dem des 20. und 21. Jahrhunderts der *Computer*: ein magischer Apparat, dessen bloße Erwähnung die Hoffnung weckt, daß zeitraubende Aufgaben sich automatisieren und auf Knopfdruck erledigen ließen. Was dabei immer wieder übersehen wird ist, daß es zwei prinzipiell verschiedene Gründe dafür gibt, *warum* Aufgaben sich zeitraubend gestalten.

Der erste liegt in der schieren Menge der zu bewältigenden Fakten. Niemand wird für das Zählen einer Hand voll Erbsen einen Taschenrechner verwenden. Wenn es dagegen um das Auswerten komplexer statistischer Daten, um die Ermittlung der Anzahl der Erwähnungen Napoleons in Tolstois *Krieg und Frieden* oder die Ermittlung von Deckungslücken in der Rentenversicherung geht, so lohnt sich der Einsatz des Computers ganz entschieden. In diesen Fällen kommt man mit einem Rechner ganz einfach schneller zu einem eindeutigen Ergebnis. Aufgaben können andererseits aber auch deshalb zeitraubend geraten, weil sie - ganz unabhängig von der zu bewältigenden Quantität und Komplexität von Daten - schlichtweg schwierig sind und sich dagegen sperren, auf eine Formel reduziert zu werden. Und das genau ist die eigentliche Pointe von Swifts Satire: Das Projekt der *Literary Engine*, so teilt uns Lemuel Gulliver mit, zielt nämlich auf nicht weniger als das ‚improving (of) speculative knowledge by practical mechanical operations‘. Erweitert werden soll wohlgerneht ein Wissen, das als *spekulativ* und gerade nicht als *empirisch* bezeichnet wird: also nicht ein Wissen, das man durch das bloße Registrieren und Zählen von Fakten erwirbt, sondern ein synthetisches Wissen, wie es erst über Prozeduren der Verknüpfung und Auslegung von Fakten gewonnen werden kann. Die eigentliche Absurdität des lagadonischen Projekts liegt deshalb für Swift in seiner Prämisse. Sie lautet, daß selbst das höherrangige spekulative Wissen endlich sei. Genug Zeit und entsprechende Apparaturen wie Institutionen vorausgesetzt, so glaubt der lagadonische Professor, könne man das Wissen einfach kombinatorisch ausschöpfen. Aber Gulliver

bemerkt sehr wohl, daß die hermeneutische Intelligenz, die zur Scheidung des Wissenswerten vom Sinnlosen gefordert ist, überhaupt nicht in der Maschine und ihren Zahnrädern, Klötzen und Seilen angelegt ist. Sie bleibt vielmehr das Privileg ausgerechnet jener willigen Studenten, die die mechanisch generierten Zeichenfolgen lesen und interpretieren müssen. Die Kurbeln drehen lassen könnte man auch von Affen - aber lesen und Sinnfragmente identifizieren müssen allemal noch Menschen. Das aber kostet Zeit.

Schon das auf Leibniz gemünzte Beispiel der Swiftschen *Literary Engine* demonstriert treffend, was CP *nicht* ist und auch nicht sein kann. CP

- zielt *nicht* auf die automatische Generierung von Bedeutungen oder Interpretationen;
- basiert *nicht* auf dem Axiom, daß Texte ein endliches, rein mechanisch ausbeutbares Bedeutungspotential enthielten;
- geht konzeptionell *nicht* in der Anwendung irgendeiner - oder mehrerer - Techniken auf,
- macht schließlich die Philologie *nicht* ‚schneller‘ oder effizienter.

Die Auskunft, was CP *nicht* ist, wird jedoch kaum befriedigen. CP ist eine relativ junge - ja, da fängt das Problem schon an: was denn nun? Technik? Methode? Disziplin? Die Frage ist nicht leicht zu beantworten, und mit einer bloßen Lexikondefinition läßt sie sich auch nicht aus der Welt schaffen. Ich werde deshalb im Folgenden besser von einer vorläufigen Bestimmung ausgehen und dann einen historischen Abriß versuchen, der die schrittweise Entwicklung dieses neuen Forschungsfeldes aufzeigen soll. Vor diesem Hintergrund werden wir im dritten Abschnitt ein Zwischenresümée ziehen, um uns Klarheit zu verschaffen über die methodischen Voraussetzungen und den Status der unterschiedlichen Praktiken und Resultate, die das Forschungsfeld CP konstituieren. Im vierten Teil des Artikels wird ein Beispiel für computerphilologische Arbeitstechniken und Projekte vorgestellt werden. An ihm läßt sich der methodische Anspruch verdeutlichen, den eine ambitionierte CP zu erheben versucht; ein Anspruch, der weit über den eines reinen Tools reicht. Wir schließen mit einem schematischen Überblick über die Komponenten und Prozesse, die das Projekt einer für die Literaturwissenschaften insgesamt relevanten CP charakterisieren.<sup>[3]</sup>

### 1. Computerphilologie: eine vorläufige Definition

Gehen wir also zunächst von einer vorläufigen und bewußt recht allgemein gehaltenen Bestimmung von CP aus. Sie lautet: ‚Computerphilologie (CP) ist ein Teilbereich des Arbeitsfeldes Humanities Computing. Die CP befaßt sich mit Aspekten der rechnergestützten Bearbeitung traditioneller philologischer Gegenstände.‘ Drei Kriterien definieren hier die CP:

- die Gegenstandsklasse, mit der sie sich befaßt: *traditionelle*

*philologische Gegenstände;*

· die spezifischen Techniken, die sie dabei verwendet:  
*rechnergestützte Verfahren;*

· der weitere methodische Kontext, in den sie als wissenschaftliche Praxis integriert ist: das *Forschungsfeld Humanities Computing*.

Auf diese drei Aspekte soll nun genauer eingegangen werden.

### 1.1 Gegenstandsbezogene Definition

Was sind *traditionelle philologische Gegenstände*? Die Problematik dieser Formulierung steckt erstens im Adjektiv *traditionell*, das man leicht im normativen Sinne mißverstehen kann, zweitens im Begriff des *Gegenstandes*, der dazu einlädt, im engsten Wortsinne an ein Buch zu denken. Nach dieser - falschen - Auffassung versteht manch einer unter traditionellen philologischen Gegenständen dann jene 50 oder 100 Bücher, von denen die Feuilletons in schöner Regelmäßigkeit immer kurz vor den Büchermessen behaupten, daß erst diese dem Zeitgenossen die rechte kulturelle Bodenhaftung verliehen.

Philologie ist jedoch keine Sache der Verbreitung solcherart vermeintlicher Gewißheiten über die materiale Konstitution ihres Gegenstandsfeldes. Wie jede Wissenschaft ist Literaturwissenschaft vielmehr eine Angelegenheit der Formulierung und Behandlung von intelligenten Fragen, die die Gegenstände in deren Bedeutung für uns selbst besser verstehen helfen. Diese Gegenstände sind in unserem Fall nun nicht nur Bücher, sondern alle Formen, in denen verschriftlichte Sprache erscheinen kann, mithin *Texte*. Der Textbegriff ist dabei nicht essentiell oder material bestimmt: es ist egal, ob ein Text als Buch, als Zeitungsartikel, als Website u.ä.m. erscheint, solange er nur eine symbolische Repräsentation von sprachlichen Äußerungen liefert. Mit anderen Worten, Texte sind definiert über ihre semiotische Funktion als sprachliche Bedeutungsträger. *Gegenstand der Philologie* aber sind damit in allererster Linie die *Fragen, die sich auf Texte* beziehen: Fragen also, die die Genese, Authentizität, Bedingungshaftigkeit, Regularität und Funktionalität von Texten als Bedeutungsträgern betreffen.

Fragen nun werden *traditionsbildend*, indem sie zur Entwicklung von Methoden stimulieren. So gibt es zum Beispiel in den Textwissenschaften Methoden, mit der Fragen hinsichtlich der Authentizität von Texten bearbeitet werden können, darunter etwa die der Stilanalyse oder der Oeuvregeschichte. Andere Methoden richten sich primär auf die Erforschung von Strukturmerkmalen von Subklassen von Texten; als Beispiel sei die Narratologie erwähnt, die sich speziell der Teilmenge der erzählenden Texte widmet. Fraglos die bekannteste und wichtigste philologische Methode aber ist die Hermeneutik, die die geregelte Explikation von Textbedeutungen zum Gegenstand hat. Wenn es in unserer vorläufigen Definition also hieß, daß die CP sich mit Aspekten der rechnergestützten Bearbeitung traditioneller philologischer Gegenstände befaßt, dann meint *traditionelle philologische Gegenstände* hier soviel wie: *textbezogene Fragen, zu deren Behandlung spezielle philologische Methoden entwickelt*

worden sind.

### 1.2 Methodenbezogene Definition

Das zweite Kriterium in unserer vorläufigen Definition ist das der *rechnergestützten Bearbeitung*. Auch hier scheint es zunächst, als liefe das auf ein ganz handfestes Merkmal hinaus; im Umkehrschluß könnte man geradezu folgern, daß keine CP stattfinden kann, wo kein Rechner im Raum ist. Das stimmt und stimmt doch wiederum nicht, denn Rechnen kann man natürlich auch im Kopf. So wie *Buch* und *Text* zwei verschiedenen Kategorien angehören - das Buch gehört zur Klasse der materiellen Entitäten, der Text hingegen zur Klasse der funktionalen - so darf auch der Begriff des *Rechners* nicht allein im konkreten Sinne verstanden werden. Was am Rechner zählt, ist nicht das, was *in* ihm zählt, sondern *daß* und *wie* er zählt: also die Tatsache, daß er *digitale Repräsentationen* der Welt nach festen Regeln verarbeitet.

Und damit läuft nun doch schon unsere vorsichtige und vorläufige Bestimmung von CP schnurstracks auf die Kernfrage hinaus: denn was markierte eine entschiedenere Differenz zur digitalen Repräsentation von Welt als die symbolische, die der natürlichsprachliche Text leistet? In der Tat ist die Vermittlung zwischen digitalen und symbolischen Repräsentationen methodisch wie philosophisch gesehen der Dreh- und Angelpunkt aller computerphilologischen Praxis. Dabei müssen wir jedoch berücksichtigen, daß das, was in der Praxis der CP im Rechner in digitaler bzw. numerischer Form repräsentiert wird, nicht die (fiktive oder reale) Welt ist, die schon der Text sprachlich repräsentiert: es ist vielmehr *der Text selbst*, der hier seinerseits zu einem in einer neuen, formaleren Symbolsprache repräsentierten System wird. Er erscheint nun nicht länger in natürlichsprachlicher, sondern in digitaler Form und kann somit algorithmengesteuert untersucht werden. Kurz, die rechnergestützte Bearbeitung traditioneller philologischer Gegenstände ist eigentlich eine Methode zur Untersuchung der *Repräsentation einer Repräsentation*. Weniger kompliziert formuliert könnte man sagen: die traditionellen Verfahren der Literaturwissenschaften sind Methoden zur Untersuchung von Fragestellungen, die sich direkt auf *Texte* beziehen. CP als Forschungspraxis entwickelt und verwendet dagegen Verfahren zur Herstellung, Analyse und Modellierung von *Textdaten*.

### 1.3 CP und Humanities Computing

In der Praxis nun erweist sich der Weg, der von empirischen Texten zu abstrakten Textdaten führt, als umso komplizierter und voraussetzungsreicher, je stärker sich das philologische Forschungsinteresse auf die Explikation von Bedeutungen richtet. An der Entwicklungsgeschichte des umfassenderen Forschungsfelds *Humanities Computing*, in das die CP einzuordnen ist, läßt sich dabei rekonstruieren, wie sich die Problemlage schrittweise von einer quantitativen zu einer qualitativen verändert hat. Für diesen Wandel sind zwei Faktoren ausschlaggebend gewesen: an erster Stelle naturgemäß der rasante Fortschritt in der informatischen Theoriebildung und der Computertechnologie; zweitens aber die



zunehmende Bereitschaft von Geisteswissenschaftlern, traditionelle Fragestellungen unter Einbeziehung fachfremder Modelle und Theorien zu rekonzeptualisieren.

Wenn man über Humanities Computing spricht (nachfolgend: HC) - also über geisteswissenschaftliche Forschungsansätze, die nicht nur punktuell rechnergestützte Verfahren zur Anwendung bringen, sondern in einem grundlegenden Sinne auf formalen Konzeptualisierungen ihrer traditionellen Forschungsgegenstände beruhen - dann müßte man eigentlich nicht nur die Philologie, sondern z.B. auch die Geschichte, die Kunstgeschichte, die Musikwissenschaft usw. berücksichtigen. Wir werden uns jedoch im anschließenden historischen Abriss auf den Aspekt *Literatur* und damit auf den Teilbereich des sog. *Literary Computing* konzentrieren, das im Kontext des HC von Anfang an eine zentrale Stellung eingenommen hat. In der Frühphase des HC erweist sich dabei allerdings die Abgrenzung zum komplementären Forschungsfeld *Linguistic Computing* als relativ unscharf. Während der Begriff *Linguistic Computing* unproblematisch als *Computerlinguistik* übersetzt werden darf, ist bei der Übersetzung von *Literary Computing* in den deutschen Begriff *Computerphilologie* an den analogen Bedeutungsüberschuß zu erinnern, der den anglo-amerikanischen Terminus *Literary Theory* als Bezeichnung für eine philosophisch ambitionierte Theoriedebatte am Beispiel literarischer Repräsentationen von dem nüchterneren und präziser, weil primär anwendungsorientiert gefaßten deutschen Begriff *Literaturtheorie* unterscheidet.

## 2. Die Entwicklung von Humanities Computing (1949 – 2002)<sup>[4]</sup>

### 2.1 Die Pionierphase 1949-1970

#### *Rechnergestützte Indices und Konkordanzen*

Mit dem Namen von Roberto Busa verbindet sich heute für uns das erste Projekt genuin geisteswissenschaftlicher Computeranwendung. Der Jesuitenpater ist deshalb auch der Namenspatron des sog. *Busa-Award*, den die federführende *Association for Literary and Linguistic Computing* alle zwei Jahre für herausragende Leistungen im Feld des HC vergibt.

Roberto Busa begann 1949 mit der Arbeit an einer Konkordanz zu den Werken Thomas von Aquins. Was ist eine Konkordanz? Fangen wir anders herum an: was ist ein Index? Ein Index ist ein sortiertes Verzeichnis, in dem alle Fundstellen für alle Wörter eines Textes angegeben werden, wobei die Fundstellen in der Regel als Kombination von Seiten- und Zeilenzahl definiert sind. Eine Konkordanz hingegen ist ein Verzeichnis, das die Fundstellen von bestimmten ausgewählten Schlüsselwörtern auflistet und dabei außerdem den Kontext zitiert, innerhalb dessen das Suchwort jeweils erscheint.

Busas Doktorarbeit nun war der Untersuchung des theologischen Konzepts der ‚Präsenz‘ in Aquins Werken gewidmet. Er erkannte dabei schon bald, daß es nicht ausreichen konnte, die

Vorkommnisse der offensichtlich relevanten lateinischen Worte *praesens* und *praesentia* zu verzeichnen, sondern daß man auch die der Präposition *in* erfassen müßte - und die kommt, wie man sich unschwer vorstellen kann, in einem lateinischen Korpus von 10,6 Millionen Wörtern nicht eben selten vor. Busa beschrieb 10 000 Karteikarten per Hand, bevor er zu der Einsicht gelangte, daß deren schließliche Auswertung, auf die das Projekt ja abzielte, ohne maschinelle Hilfe schlechterdings nicht mehr zu leisten sein würde. Die Karteikarten wurden daher mühsam auf Lochkarten übertragen, und ein Rechner generierte dann einen Index. 1974 erschien der *Index Thomisticus* gedruckt in 31 Bänden mit insgesamt 36 000 Seiten - das Nettoresultat der Bemühungen von nunmehr insgesamt 66 Forschern, die über 5 Jahre zusammen eine Million *man-hours* investiert hatten! Busas Pionierleistung wurde später als CD-ROM mit dem Antike und Moderne treffend verklammernden lateinischen Titel *Thomae Aquinatis Opera Omnia cum hypertextibus in CD-ROM* publiziert.<sup>[5]</sup>

Auch nach Busa sollte die Avantgarde des HC zunächst hauptsächlich aus Bibelforschern und Theologen bestehen. Dazu zählt etwa John William Ellison, der 1957 an der Universität Harvard mit Erfolg seine zweibändige Dissertation *The use of electronic computers in the study of the Greek New Testament text* verteidigte. 1959 schließlich wurde mit der von Stephen Maxwell Parish herausgegebenen *Cornell Concordance to the poetry of Matthew Arnold* erstmals eine computerphilologische Studie publiziert, die einem säkularen Text galt. Das Buch bestand aus der photographischen Reproduktion eines Computer-Printouts von 995 Seiten Umfang, dessen Erstdruck allein stolze 38 Stunden Rechnerzeit in Anspruch genommen hatte.

#### *Stylometrie und Autorschaftsattribuion*

Einen neuen Anwendungsaspekt eröffnete das 1961 erschienene Buch von MacGregor und Morton, *The Structure of the Fourth Gospel*. Die beiden Forscher hatten einen Computer eingesetzt, um die Länge von Sätzen und Absätzen im Johannesevangelium zu analysieren, dessen Zuschreibung an Johannes als problematisch galt. Beide kamen - allerdings auf der Grundlage völlig unterschiedlicher Interpretationen der so generierten Daten - zu der Schlußfolgerung, daß das 4. Evangelium zwei verschiedene Quellen haben mußte. Damit wurde erstmals in die Praxis umgesetzt, was der Londoner Mathematikprofessor Augustus de Morgan bereits 1851 vorgeschlagen hatte, nämlich die Bestimmung der Autorschaft von Texten anhand der Untersuchung von Wortlängen. Autorschaftsattribuion ist heute weiterhin eine der Zielsetzungen, die man mit der sogenannten *Stylometrie* verfolgt, u.a. in der forensischen Linguistik.

Andrew Morton übrigens erlangte schon bald Berühmtheit für die mitunter recht kühnen Thesen, die er aus seinen computergestützten stylometrischen Untersuchungen ableitete. 1963 etwa behauptete er in der *New York Times*, mit einer statistischen Analyse von Satzlängen und Funktionswörtern bewiesen zu haben, daß der Apostel Paulus nur eine der vier ihm zugeschriebenen Episteln selbst verfaßt haben könne.

John Ellison, ein anderer Bibelforscher, konterte daraufhin in seinem gewitzten Artikel ‚Computers and the Testaments‘, indem er Mortons Methode auf zwei moderne Textkorpora anwandte. Das Resultat war einigermaßen verblüffend, denn dabei kam folgendes heraus: erstens, James Joyce ist keinesfalls der Autor des *Ulyssess*; zweitens aber und noch viel erstaunlicher: Andrew Morton ist nicht der Autor der Texte Andrew Mortons!

Abgesehen von dieser mit Esprit geführten methodologischen Kontroverse stellt sich die Frühphase des HC im Rückblick nicht zuletzt als eine Material- und Ressourcenschlacht von einem Ausmaß dar, das Swifts lagadonische *Literary Engine* spielend in den Schatten stellt. Morton selbst erinnerte sich 1980:

Memories of the early days are all of paper tape. It waded in and out of every machine, it dried and then cracked and split or it got damp when it lay limp and then sagged and stretched. Sometimes it curled round you like a hungry anaconda, at others it lay flat and lifeless and would not wind. Above all it extended to infinity in all directions. A Greek New Testament, half a million characters, ran to a mile of paper tape, and the complete concordance of it ran to seven miles.<sup>[6]</sup>

### *Computerlinguistik*

Gegen Ende 60er Jahre erschienen schließlich die ersten deutschsprachigen Studien. Sie stammten in der Regel von Linguisten, so u.a. Winfried Lenders *Untersuchungen zur automatischen Indizierung mittelhochdeutscher Texte* (1969) und Peter Schmidts *Der Wortschatz von Goethes ‚Iphigenie‘: Analyse der Werk- und Personensprache mit EDV-Hilfe; mit Wortindex, Häufigkeitwörterbuch und Wortgruppentabellen* (1970). In den 70er Jahren etablierte sich die Computerlinguistik dann schnell als eigenständiger Arbeitsbereich und bald auch schon als Lehrfach. Sie ging dabei im Kern aus einer Forschungsrichtung hervor, für die Sally und Walter Sedelow schon 1965 den Begriff *Computational Stylistics* geprägt hatten. Die besondere Bedeutung, die gerade im deutschsprachigen Raum der Computerlinguistik bei der rechnergestützten Textanalyse zukommt, verdankt sich nicht zuletzt einer Eigenart unserer Sprache, nämlich ihrem Flexionsreichtum. Anders als im Englischen sind mechanisch generierte Wortlisten und Indices deutscher Texte für uns von relativ geringer Aussagekraft, solange darin jede Wortform als Einzeleintrag erscheint (*mein, meine, meines, meiner, meinem* usw.). Bei der Vielzahl der Präfixe und Endungen - von der Varianz des Wortstammes etwa bei manchen Verben ganz zu schweigen - ergibt sich schnell die Notwendigkeit zur Rückführung von *tokens* auf *types*, d.h. von konkreten Wörtern auf einen gemeinsamen Wortstamm (Lemma), wenn nicht gar auf ein abstraktes semantisches Konzept.

## 2.2. Tools, Textarchive, Institutionalisierung

### *Erste Tools: Parser, Concordancer und Kollationsprogramme*

Anfang der siebziger Jahre veröffentlichte Irmtraut Barth zwei deutlich technisch orientierte Bücher, deren Titel zwar den Konnex zur Philologie bewußt hielten, dabei jedoch auf diese linguistische Problematik hinwiesen, der man in der Pionierphase des HC recht wenig Aufmerksamkeit gewidmet hatte: *Philologische Texterschliessung. 1. Der Zerlegprogrammgenerator ZPG des Anwendungssystems PHILTEX* (1970) und *Philologische Texterschliessung. 2. Der Häufigkeitsprogrammgenerator HPG des Anwendungssystems PHILTEX für die Datenverarbeitungsanlage IBM 7094* (1971). Die computerlinguistische Frage, die damit angesprochen wurde, läßt sich so fassen: wie kann man Sätze und Wörter vor der mechanischen Indexierung automatisch in syntaktische und semantische Einheiten gliedern? Diese Techniken der Zergliederung kennen wir heute unter den Begriffen *Parsing* und *Lemmatisierung*. Das von Barth beschriebene Programm PHILTEX war eines der computerlinguistischen Tools, mit dem im Prinzip auch Nichtinformatiker einen Text – vorausgesetzt, daß er im entsprechenden technischen Format und auf dem richtigen Medium zur Verfügung stand - bearbeiten können sollten.

Prototyp solcher bewußt als Anwendungswerkzeug konzipierten Programme war das 1967 von Donald Russel im Atlas Computer Laboratory entwickelte Produkt COCOA. Das Kürzel hat nichts mit Kokosmilch zu tun, sondern steht für *Count and Concordance Generation on Atlas* - vereinfacht gesagt also für ein Programm, das Wörter zählen und nach bestimmten Kriterien sortierte Wortlisten und Konkordanzen generieren kann. Was COCOA für Philologen ungemein attraktiv machte war die Tatsache, daß das Programm - wenn auch in einem sehr bescheidenen Umfang - modifizierbar war. COCOA erlaubte es nämlich, eine Sequenz aus bis zu drei Buchstaben als neue Einheit zu definieren und so dem vorgegebenen ASCII-Maschinenalphabet neue Einträge hinzuzufügen, was etwa die Bearbeitung von Texten mit Umlauten oder anderen Sonderzeichen ermöglichte.

### *Aufbau von Textarchiven*

Die Entwicklung von Tools war der eine Faktor, der zumindest in der anglo-amerikanischen Welt dem HC in den späten 60er Jahren zum Durchbruch verhalf. Die Zahl der Forschungsprojekte stieg nun sprunghaft an; 1965 etwa wurden alleine 50 rechnergestützte Untersuchungen zum Werk Shakespeares erfaßt. Das gewachsene Interesse und der Bedarf zum wissenschaftlichen Austausch fanden ihren Niederschlag in der Gründung der bis heute tonangebenden Zeitschrift *Computers and the Humanities* (CHUM) im Jahr 1966. Der zweite wesentliche Faktor für die Konstituierung des Forschungsfelds aber war die Etablierung von Forschungszentren und elektronischen

Textarchiven, damit also der Beginn der Institutionalisierung von HC. Hier spielte die Katholische Universität Louvain die Vorreiterrolle, an der 1968 das *Centre for the Electronic Processing of Documents* (CETEDOC) unter der Leitung von Paul Tombeur gegründet wurde.<sup>[7]</sup> Das CETEDOC produzierte u.a. die *Library of Christian Latin Texts* und das *Archive of Celtic-Latin Literature*.

HC-Zentren und elektronische Textarchive entstanden nun in rascher Folge, und zwar vornehmlich im englischsprachigen Raum. 1969 gründete die American Philological Society am Dartmouth College in New Hampshire das *Repository of Greek and Latin machine-readable texts*; im gleichen Jahr wurde in Cambridge von Roy Wisbey das *Centre for Literary and Linguistic Computing* etabliert. Hier fand dann auch im folgenden Jahr das erste *Symposium on Literary and Linguistic Computing* statt. 1972 schließlich begann an der University of California, Irvine die Arbeit am *Thesaurus Linguae Graeca*. Dieses Projekt sollte ein Vierteljahrhundert in Anspruch nehmen; am Ende stand ein Korpus in elektronischer Form zur Verfügung, der über 8200 griechische Texte von beinahe 3000 Autoren umfaßt: ein gigantischer Thesaurus von 69 Millionen Wörtern Umfang, der auf CD-ROM gespeichert ist. In England begründete 1976 Lou Burnard das *Oxford Text Archive*, das sich schnell zu einem der größten elektronischen Textarchive moderner Sprachen entwickelte. Viele dieser frühen Archive verfahren nach einer Maxime, die Burnard als ‚the dustbin policy of archiving‘<sup>[8]</sup> bezeichnet hat: gesammelt wurde alles, in unterschiedlichen Formaten und Codierungen. Standards gab es noch nicht - wie konnte es das auch in einem Entwicklungsstadium, wo sich die Geister manchmal noch an Fragen schieden wie der, ob man überhaupt die Unterscheidung in Groß- und Kleinschreibung erfassen können mußte, wenn viele Systeme ohnehin nur Kleinbuchstaben anzeigen konnten?

Was die Entwicklung von Tools anging, so traten Anfang der 70er Jahre die ersten Kollationsprogramme auf den Plan. Ihre Funktionalität überstieg die der alten Konkordanzprogramme erheblich. Zwar wurden weiterhin im ersten Schritt Wortlisten und Konkordanzen generiert - dies aber nicht nur für einen, sondern gleich parallel für mehrere Texte. Kollationsprogramme dienen nämlich hauptsächlich dazu, verschiedene Versionen eines Textes zu vergleichen, Varianten von Textpassagen hervorzuheben und dem jeweiligen Basistext die entsprechenden Partien des Referenztextes zuzuordnen. Ein noch heute verwendetes, laufend weiterentwickeltes Kollationsprogramm aus dieser Phase heißt schlicht COLLATE. In der aktuellen Version kann COLLATE bis zu 200 (!) Texte kollationieren. Bekannt geworden ist daneben unter dem Kürzel OCP auch das *Oxford Concordance Programme*. Schon die nüchternen Bezeichnungen der beiden Tools kennzeichnen den neuen Pragmatismus, der jetzt, wo der Forschung zunehmend Texte im elektronischen Format zur Verfügung standen, den experimentellen Gestus der Pionierphase abzulösen begann. Dem ursprünglichen Charakter des Unternehmens HC und der Vorliebe seiner Adepten für selbstironische Namensgebungen gerecht wurde da noch weit eher die Software OCCULT, mit der George Petty and William Gibson 1970 Aufmerksamkeit erregten. OCCULT war dabei

alles andere als ein Fall für Harry Potter: hinter dem Kürzel verbarg sich die *Ordered Computer Collation of Unprepared Literary Text*.

#### *Institutionalisierung: Verbände, Zeitschriften, HUMANIST*

Zwei Jahre nach dem von Roy Wisbey veranstalteten ersten Literary Computing-Symposium fand 1972 das zweite Symposium über den Einsatz von Computern in Literaturwissenschaften und Linguistik in Edinburgh statt. Dabei wurde die *Association for Literary and Linguistic Computing* (ALLC) gegründet, deren offizielle Verbandszeitschrift *Journal of Literary and Linguistic Computing* (JLLC) seitdem bei der Oxford University Press erscheint.<sup>[9]</sup> Zusammen mit der Zeitschrift *Computing in the Humanities* (CHUM) und dem stärker computerlinguistisch orientierten nordamerikanischen Schwesterverband *Association for Computing in the Humanities* (ACH) repräsentieren damit seit nunmehr dreißig Jahren zwei internationale Verbände und Fachzeitschriften das Forschungsfeld. Hinzu trat 1987 das von Willard McCarty ins Leben gerufene Diskussionsforum HUMANIST, das ein Muß für jeden Computerphilologen ist und über ein komplettes Archiv aller Diskussionsbeiträge und Anfragen aus mehr als 16 Jahren verfügt.<sup>[10]</sup> Seit den Gründerjahren oszilliert die internationale Debatte in HUMANIST zwischen der Diskussion methodologischer und philosophischer Grundsatzfragen einerseits, andererseits aber der Erörterung technologischer Neuerungen und aktueller Forschungsprojekte.

#### *Erste methodische Selbstreflexionen*

In den Diskussionen, die man Dank dieser formellen Institutionalisierung von HC in ihrer Abfolge und Entwicklung genauer rekonstruieren kann, setzte ab Mitte der 70er Jahre erstmals die bislang vernachlässigte methodische Selbstreflexion des neuen Forschungsfeldes ein. Zwei große Themen standen hierbei im Mittelpunkt: erstens die Frage, inwieweit durch das elektronische Medium eine Veränderung des Textbegriffes selbst notwendig werden würde; zweitens die Frage, ob HC einen Anspruch auf epistemologischen und methodologischen Eigenwert erheben konnte. Zwei Zitate mögen dies verdeutlichen. Paul Tombeur, der Leiter des CETEDOC an der Université catholique de Louvain, unterstrich den methodologischen Aspekt 1972 so:

One important advantage of the application of data-processing methods lies in the complete analysis of the problems. The necessity of applying rigorous logic to all the stages of a study, of breaking down each stage into many simple elements, gives paramount importance to the methodological approaches to our problems. The computer forces us to master our problems as perfectly as possible; otherwise we run the risk of being furnished with deceptive output, and of having our principal questions left unanswered.<sup>[11]</sup>

Roberto Busa hingegen hob 1975 auf den Textbegriff und den epistemologischen Aspekt ab:

Electronic data processing marked the beginning of a new era in the

transfer of human information...At Gutenberg's time typesetting started a new era in the distribution of human knowledge. Today, we have made another jump: we are now able to use an electronic alphabet which can be processed by machine at 'electronic' speeds and distances. But we are still at the starting point of the new era as far as language processing is concerned...[the computer] is just a tool, an off-line continuation of man's fingers, as fingers may be described as the bodily computer on-line to the mind. It is man's ingenuity which has to feed data and programs into it. In language processing the use of computers is not aimed towards less human effort, or for doing things faster and with less labour, but for more human work, more mental effort; we must strive to know more systematically, deeper, and better, what is in our mouth at every moment, the mysterious world of our words. (ALLC B 4, 1975:1)[\[12\]](#)

Die Akzentsetzungen unterscheiden sich: während Tombeur eine vollständige Durchdringung der behandelten Probleme erhofft und die Verpflichtung zu rigoroser Stringenz und Perfektion betont, die durch den Computer erzwungen werden, hebt Busa hervor, daß der Computer weiterhin nur ein Werkzeug sei, das zwar eine neue Ära in der Wissensvermittlung eingeläutet hat, den Menschen aber letztlich nicht weniger, sondern intensivere Anstrengungen um die behandelten Gegenstände abverlangen wird. Tombeur wie Busa reagierten hier beide auf das immer stärker verspürte Bedürfnis, eine theoretische Legitimation für die Praxis des HC nachzuliefern, und beide wiesen dabei entschieden darauf hin, daß HC keinesfalls eine Sache des *bigger, faster, further, more* sein könne, sondern eine des *better* sein müsse: also eines qualitativen und nicht eines quantitativen Zuwachses. Solcherart Positionsbestimmung wurde Mitte der 70er Jahre nicht zuletzt deshalb unumgänglich, weil HC nunmehr auch im Bereich der Lehre auf den Plan trat. So hielt Susan Hockey bereits 1976 an der Oxford University eine Reihe von Vorlesungen zum Thema *Computing in the Arts*, die Geisteswissenschaftler nicht allein mit Methoden und Anwendungen vertraut machen wollten, sondern auch einen Crash-Kurs in der Programmiersprache SNOBAL beinhalteten. Horribile dictu: der Philologe als Programmierer?

#### *Die Technische Revolution der 80er: PC und Internet*

Schon bald allerdings erwies sich, daß es nicht die gerade erst eröffnete methodologische Debatte, sondern eine handfeste technische Revolution war, die der weiteren Entwicklung des Forschungsfeldes im Laufe der 80er Jahre eine vollkommen neue Qualität und Dynamik verlieh. Man kann das unschwer auf zwei Stichworte bringen: PC und Internet. Mit der sich rasant durchsetzenden Dezentralisierung von Rechnersystemen, die den Geisteswissenschaftlern in Gestalt von PCs und Macs anpassungsfähigere und leichter zu handhabende Werkzeuge zur Verfügung stellten als die klassischen Rechenzentren, zugleich aber eine internationale Vernetzung von Forschungsinitiativen ermöglichten, stellte sich umso dringlicher die Frage nach der Vereinbarung von verbindlichen Standards. Dabei ging es nicht länger darum, in welchem Zeichencode und auf welchen Medien Texte elektronisch gespeichert werden sollten; der ASCII-Zeichencode hat sich schon lange durchgesetzt und die Frage des Speichermediums sollte Dank der technischen Entwicklung binnen weniger Jahre so gut wie obsolet werden. In den Vordergrund rückte jetzt vielmehr das Desiderat einer

Standardisierung der *Auszeichnung* von elektronischen Texten. Auszeichnung heißt in diesem Zusammenhang: die explizite und eindeutige Beschreibung von relevanten formalen und inhaltlichen Textmerkmalen. Erst diese Explikation nämlich macht aus Texten genuine *Textdaten*. Die Methode dieser elektronischen Textauszeichnung nennt man *tagging* (ein übertragener Gebrauch des Englischen *tag*, was soviel wie *Preisschild* oder *Anhänger* heißt). Auch das Produkt des Prozesses wird mit einem Wort bezeichnet, das man eigentlich aus dem Handel kennt: *Mark-up*. Der Begriff bedeutet im Literalsinn soviel wie Aufschlag, Mehrwert oder Gewinnspanne. Die ökonomische Metapher ist durchaus passend, denn auch für den Geisteswissenschaftler erhält ein Text, der sich durch *tagging* in *Text+Mark-up = Textdaten* verwandelt, einen ganz entscheidenden Mehrwert. Aber wie in der ökonomischen so schlägt sich auch in der wissenschaftlichen Praxis Mehrwert nur dann in Tauschwert nieder, wenn man sich auf eine gemeinsame Währung einigen kann.



### 2.3. Etablierung von Standards und Kommerzialisierung elektronischer Texte

#### *SGML und TEI*

1986 wurde deshalb die sog. *Standard Generalized Markup Language* (SGML) als erster internationaler Markup-Standard für elektronische Texte definiert. SGML diente und dient allerdings im wesentlichen der formalen und typographischen Textauszeichnung; die für Literaturwissenschaftler relevanten grammatischen, syntaktischen, semantischen und rhetorischen Kategorien werden dabei nicht berücksichtigt. Bereits 1987 etablierten daher die drei führenden Verbänden ALLC, ACH und die *Association for Computational Linguistics* (ACL) die *Text Encoding Initiative* (TEI). Ziel dieses internationalen Joint Ventures war es, Richtlinien für eine standardisierte Textauszeichnung in den Geisteswissenschaften auf der Grundlage von SGML zu erarbeiten. Die erste Version der Richtlinien erschien 1992, die bislang vierte und letzte Version im Jahr 2003 - ein Mammutwerk von nunmehr fast 1000 Seiten! Mittlerweile gibt es mit TEI Lite eine nutzerfreundlichere Kurzfassung dieser Textauszeichnungsbibel. Daneben hat man mit XML (Extendable Markup Language) auch noch eine Metasprache geschaffen, mit der geregelt wird, wie Textauszeichnungssprachen prinzipiell strukturiert und speziellen Auszeichnungsbedürfnissen angepaßt werden können.<sup>[13]</sup> Fast alle elektronischen Textarchive der Welt haben mittlerweile den TEI-Auszeichnungsstandard übernommen.

Eines der ersten Projekte, in dem die TEI-Richtlinien konsequent angewendet wurden, begann 1988 an der Brown University in den USA: das *Women Writers Project*, das größte elektronische Textarchiv mit Werken prä-viktorianischer englischsprachiger Autorinnen. Alle Texte in diesem Archiv sind TEI-konform ausgezeichnet. Das gilt auch für das 1991 gegründete *British National Corpus Project* (BNCP), in dem bislang über 4000 moderne britische Texte in SGML / TEI erfaßt worden sind. Das BNCP war zugleich eines der ersten Projekte, in dem akademische Forschungszentren und kommerzielle Verlage in großem Stil kooperierten. Die konsequent TEI-konforme Textauszeichnung ganzer Textkorpora erfordert nämlich einen Aufwand, der den für die bloße Erfassung von Texten im elektronischen Format um ein Vielfaches übersteigt. Verdienstvolle nonkommerzielle Initiativen wie das noch heute existierende *Projekt Gutenberg*, in dem bereits seit den 80er Jahren kollaborativ Texte im reinen ASCII-Format gesammelt und kostenfrei zur Verfügung gestellt wurden, konnten dies zumeist nicht leisten. 1993 - also im selben Jahr, als das *National Centre for Supercomputing Applications* in den USA mit MOSAIC den ersten Webbrowser zur Verfügung stellte und damit das WWW, so wie wir es heute kennen, ins Leben gerufen wurde - veröffentlichte beispielsweise der Verlag Chadwyck-Healey die *Patrologia Latina*-Datenbank. Den seinerzeitigen Pfund/DM-

Umrechnungskurs zugrundegelegt kostete die auf etlichen CD-ROMS publizierte und damals weltweit umfangreichste Text-Datenbank in etwa soviel wie eine Luxuslimousine: 27 000 Pfund.<sup>[14]</sup> Das war für philologisch anspruchsvolle, sorgfältig erstellte und mit textkritischem Mark-up versehene elektronische Texteditionen durchaus kein außergewöhnlicher Preis. Als Chadwyck-Healey drei Jahre später erstmals *Goethes Werke* in der Sophienausgabe im elektronischen Format auf den Markt brachte, erhielt ich selbst Gelegenheit, eine Vorabversion der CD-ROM zu testen. Zur Sicherheit bat ich einen Hamburger Kollegen, die wertvollen Original-CD nachts in seinem Stahlschrank zu verwahren. Prompt wurde in der darauffolgenden Nacht in das Büro meines Kollegen eingebrochen; Rechner, Scanner, Bildschirme und ganze Regale voller Software verschwanden. Mit dem elektronischen Goethe allerdings hatten die Diebe offensichtlich nichts anfangen können; die CD lag in dem ansonsten leeren Schrank. Der Stein, der mir vom Herzen fiel, hatte einiges Gewicht: *Goethes Werke* auf CD-ROM kosteten damals DM 12 000.

#### 2.4. Von der Serviceeinrichtung zum Lehrfach

Die Erschließung und standardisierte Auszeichnung neuer und zunehmend umfangreicherer Textkorpora, sofern sie nicht von kommerziellen Verlagen betrieben wurden, führte an den Universitäten zur allmählichen Trennung von reinen Textarchiven und genuinen HC-Zentren. In der Frühphase waren die HC-Zentren zumeist als Serviceeinrichtungen für Geisteswissenschaftler etabliert worden, die den universitären Rechenzentren angegliedert waren. Das konnte durchaus vorteilhaft sein, wie u.a. das Beispiel des *Center for Computing in the Humanities* (CCH) an der University of Toronto bewies: die Rechenzentren als Großabnehmer der IT-Branche verfügten oft über gute Kontakte zur Wirtschaft, die sich mitunter in der Bereitstellung von privaten Fördermitteln für technische Entwicklungsvorhaben im Bereich HC niederschlugen. Eines der noch heute interessantesten computerphilologischen Tools, das Softwarepaket TACT (*Textual Analysis Computing Tools*) ist die Frucht eines solchen mehrjährigen Kooperationsvertrags zwischen IBM Canada und der University of Toronto. Die 1992 erfolgte Gründung des *Institute for Advanced Technologies in the Humanities* (IATH) an der University of Virginia wurde ebenfalls durch eine millionenschwere Förderung durch IBM ermöglicht; die aktuelle Sponsorenliste dieses Instituts liest sich weiterhin wie das *Who is Who* der nordamerikanischen IT-Branche.<sup>[15]</sup>

Das IATH in Virginia bietet zugleich ein gutes Beispiel für den mittlerweile an führenden Universitäten sich etablierenden neuen Typus von HC-Einrichtungen, die sich eher als ambitionierte computerphilologische *Think Tanks* denn als subsidiäre Serviceeinrichtungen verstehen. In Virginia etwa ging aus dem IATH das von Johanna Drucker geleitete *Speculative Computing Laboratory* hervor, wohingegen die Sammlung und Aufbereitung elektronischer Texte an das *Electronic Text Center* (ETC) übertragen wurde, eine der Universitätsbibliothek zugehörige Einrichtung, die eines der weltweit

umfangreichsten und bestausgestatteten elektronischen Textarchive betreut.[16] Wie das IATH haben sich viele der etablierten HC-Zentren in Nordamerika, England und Europa während der letzten zehn Jahre verstärkt um eine Hinwendung zur Grundlagenforschung und um die Förderung innovativer Projekte bemüht, sich von den Rechenzentren abgekoppelt und den Aufgabenbereich der elektronischen Textedition und Archivierung an die Bibliotheken abgegeben. Grundsätzlich ist das Selbstverständnis der avancierten HC-Zentren außerhalb Deutschlands nicht länger das von elektronischen Textarchiven, Tools-Entwicklern und Serviceeinrichtungen für die Geisteswissenschaften; sie verstehen sich vielmehr als Vertreter eines HC, das tendenziell Anspruch auf den Status einer eigenständigen akademische Disziplin erheben kann.

Seit der Gründung der elektronischen Diskussionsliste HUMANIST durch Willard McCarty im Jahr 1987 ist dies der Grundtenor einer nunmehr seit fünfzehn Jahren mit Verve geführte Debatte. Eines der zentralen Anliegen ist dabei die Etablierung eines HC-Curriculums. Wie schon bei der TEI-Initiative, die ohne nennenswerte Beteiligung oder auch nur Beachtung seitens der deutschen Literaturwissenschaften erfolgte, hinken auch bei der Institutionalisierung von HC bzw. Computerphilologie und der Curricularentwicklung die deutschen Universitäten der Entwicklung bislang weit hinterher. Führend ist hier derzeit Kanada, wo erstens die Frage der Curricula für das Lehrfach HC engagiert diskutiert wird,[17] zweitens ein Sonderprogramm zur Schaffung von HC-Lehrstühlen und Forschungszentren aufgelegt und drittens entsprechende Forschungsinitiativen mit substantiellen Mitteln gefördert werden - so im Jahr 2002 mit der Bereitstellung von knapp 7 Millionen Dollar für das universitätsübergreifende Projekt TAPOR (*Textual Analysis Portal for Research*).[18] An führenden Universitäten in Großbritannien und Kanada können Studierende schon jetzt einen Bachelor und teilweise einen MA in HC erwerben; in Deutschland ist das von der Hamburger *Arbeitsstelle Computerphilologie* (ACP) betreute ‚Studienmodul CP‘ bislang die einzige im Ansatz vergleichbare Initiative.[19]

Damit wären wir im Jahre 2003 und also in der Gegenwart angelangt. Dem Leser wird nicht entgangen sein, daß zuletzt zunehmend von Initiativen im englischsprachigen Raum die Rede gewesen ist. Auch in Kontinentaleuropa ist HC durchaus an einigen Universitäten und Instituten vertreten, so etwa in Norwegen, Frankreich und Italien.[20] In Deutschland indes gibt es – wohlgermerkt von der Computerlinguistik abgesehen - bislang kaum eine vergleichbare universitäre Einrichtung. Der derzeit aktuellste internationale Überblick *Institutional Models for Humanities Computing* von Willard McCarty und Matt Kirschbaum[21] verzeichnet (abgesehen von der Hamburger Arbeitsstelle) mit dem an der Universität Köln angebotenen Fach *Historisch-kulturwissenschaftliche Informationsverarbeitung*,[22] dem *Institut für Multimedia und Datenverarbeitung in den Geisteswissenschaften* an der Universität Rostock[23] und dem *Bereich Historische Fachinformatik* an der Humboldt-Universität Berlin[24] ganze drei Initiativen, von denen allerdings keine dezidiert computerphilologischer Natur ist. Das ist umso bedauerlicher, als ausgerechnet eine der allerersten Konferenzen

zum Thema schon 1960 an der Eberhard-Karls Universität in Tübingen abgehalten wurde, nämlich das *Kolloquium über maschinelle Methoden der literarischen Analyse und der Lexikographie*, das gemeinsam von Pater Buso und IBM Deutschland organisiert wurde. 1966 schuf die Universität Tübingen zudem die erste akademische Stelle in Deutschland in diesem Bereich, und 1970 wurde dort die *Arbeitsstelle für literarische und elektronische Datenverarbeitung* eingerichtet. Unter der Leitung von Wilhelm Ott erbrachte diese Einrichtung mit der Entwicklung des Editionsprogramms TUSTEP eine international anerkannte Pionierleistungen. Die Tübinger Initiative konzentrierte sich dabei naturgemäß nahezu ausschließlich auf editionsphilologische und lexikographische Fragestellungen.<sup>[25]</sup> HC in dem umfassenderen Sinn, wie ihn dieser historische Abriss zugrundegelegt hat, bleibt in Deutschland bislang auf wenige Einzelinitiativen beschränkt, die oftmals eher der Kultur- und Medienwissenschaft als der Philologie zuzurechnen sind. Es gibt mittlerweile jedoch eine deutliche Tendenz zur Vernetzung dieser Initiativen, zu der nicht zuletzt das von Fotis Jannidis, Georg Braungart und Karl Eibl betreute Internetforum [www.computerphilologie.de](http://www.computerphilologie.de) und das daraus hervorgegangene *Jahrbuch Computerphilologie* beigetragen haben.

Die Hamburger ACP bleibt so bislang die einzige ihrer Art, insofern sie erstens eine *interdisziplinäre* Initiative darstellt, in der Lehrende aus zwei Fachbereichen kooperieren, dem Fachbereich Sprach-, Literatur und Medienwissenschaften und dem Fachbereich Informatik. Zweitens haben die Literaturwissenschaftler, Linguisten und Informatiker, die in der ACP an einem Tisch sitzen, die Frage der Gestaltung eines Curriculums als Punkt eins auf ihre Tagesordnung gesetzt und mit der Einrichtung des Hamburger Studienmoduls CP einen wesentlichen Schritt getan. Die jüngst erfolgte Einrichtung von CP-Lehrstühlen an der TU Darmstadt und der Ludwig-Maximilians Universität München wird sicherlich neue Impulse für die Einrichtung des Lehrfachs CP an deutschen Universitäten geben, die diesen ersten Ansatz fortführen können.

Womit wir bei einem gewichtigen Einwand wären: taugt HC bzw. CP denn überhaupt zu einem ‚richtigen‘ akademischen Fach? Diese Frage kann man mit einer bloßen historischen Rekonstruktion nicht beantworten. Lassen Sie uns deshalb, bevor wir zu einem praktischen Beispiel kommen, zumindest ansatzweise auch eine systematische Einordnung der CP in der Wissenschaftslandschaft versuchen.

### 3. CP in Relation zu den etablierten Disziplinen

Fotis Jannidis hat 1999 an den Anfang seiner Überlegungen zum Thema ‚Was ist CP?‘ die Abgrenzung der CP zur Computerlinguistik (CL) gestellt. Für ihn liegt der entscheidende Unterschied darin, daß die CL hauptsächlich unsere Gegenwartssprache analysiert, während die CP

‚historische Sprachstufen untersucht mit einem deutlichen Schwerpunkt im Bereich der Edition und Interpretation fiktionaler Texte.’ [26] Diese Gegenüberstellung von gesprochener Sprache und historischen Sprachformen ist allerdings insofern problematisch, als sie eigentlich zwei heterogene Kriterien miteinander kombiniert: die systematische Unterscheidung von ‚gesprochene vs. geschriebene Sprache’ auf der einen Seite, die historische von ‚aktuelle vs. historische Sprachform’ auf der anderen. Tatsächlich liegt der entscheidende Unterschied von CL und CP, den Jannidis meint, denn auch nicht im *Was*, sondern im *Wie*. Ein wesentliches Forschungsziel der CL ist seit jeher die automatisierte Disambiguierung und Lemmatisierung von sprachlichen Äußerungen, d.h. die Ermittlung der jeweils gemeinten Bedeutung eines Wortes, und die Rückführung von Wörtern auf Grundformen und semantische Konzepte. Computerlinguistische Untersuchungen hören insofern, was den Komplexitätsgrad der analysierten Textphänomene angeht, zumeist genau da auf, wo es für den weniger an der Wort- als der Textbedeutung interessierten Philologen überhaupt erst interessiert wird: nämlich beim Übergang von der Wort- und Satz- zur Textebene.

Schon das Beispiel lyrischer Texte (bei denen eine Analyse bis hinab zur Ebene der Morpheme auch und gerade für den Literaturwissenschaftler wichtig sein kann) zeigt indes, daß auch diese Abgrenzung nur relative Gültigkeit haben kann. Es ist deshalb vielleicht grundsätzlich sinnvoller, die systematische Einordnung der CP in einem weiter gefaßten Kontext und zudem mehrdimensional vorzunehmen. Hierzu hat Walter von Hahn einen Vorschlag unterbreitet, der die CP in den über ihren Gegenstand wie über ihre Methoden gestifteten Bezügen innerhalb eines Systems der Wissenschaften zu lokalisieren sucht:

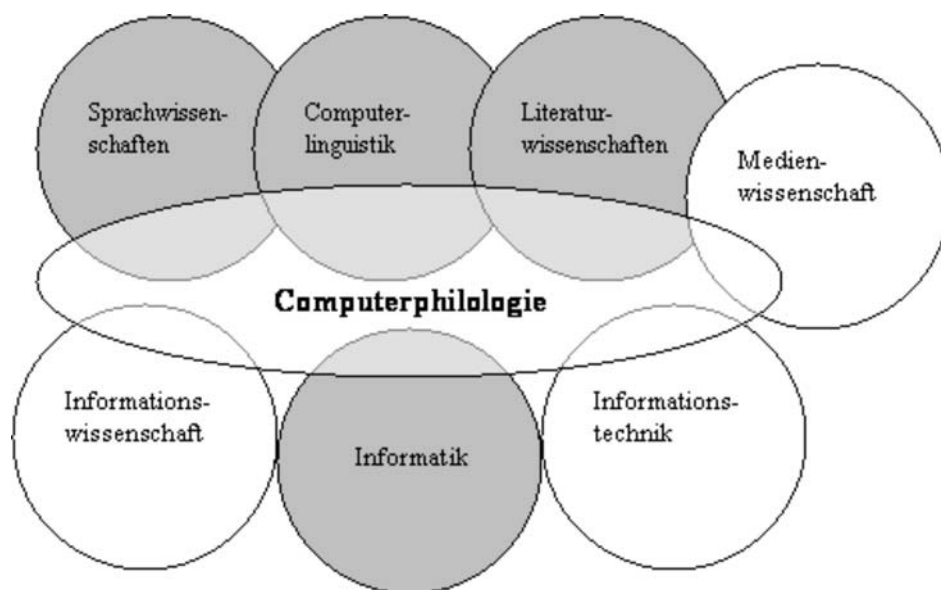


Abbildung 3: Fachliche Lokalisierung der CP nach von Hahn [27]

Naturgemäß stellt sich bei einer derartigen Kontextualisierung sofort

die bereits erwähnte Frage, ob CP denn überhaupt als gleichrangige Disziplin gewertet werden kann. Die Debatte hierüber dauert an; ihr derzeit letzter Stand ist im vierten *Jahrbuch für Computerphilologie* (2002) dokumentiert worden.<sup>[28]</sup> Aus der Wissenschaftsgeschichtsforschung wissen wir, daß die Zuerkennung des Disziplinstatus von mehreren Faktoren abhängig ist. Die beiden Wissenschaftshistoriker Martin Guntau und Hubert Laitko haben 1987 unter der Überschrift ‚Entstehung und Wesen wissenschaftlicher Disziplinen‘<sup>[29]</sup> dargelegt, daß es keine eigene Theorie des Phänomens ‚Wissenschaftsdisziplin‘ gibt, wohl aber eine Liste von Merkmalen, anhand derer man gewöhnlicherweise eine Disziplin identifiziert. Ihr Fazit:

Eine disziplinäre Gemeinschaft wird primär durch die auf den disziplinspezifischen Gegenstand gerichtete Erkenntnisintention und -disposition der Wissenschaftler zusammengehalten, sekundär durch institutionelle Formen, die diese Gemeinschaft hervorbringt bzw. in denen sie sich bewegt. Sie besteht aus Individuen, die bezogen auf den Gegenstand der Disziplin kompetent kommunizieren.<sup>[30]</sup>

Institutionalisierung als nachgeordnete Bedingung der Disziplingenese sichert zwar die notwendige Permanenz einer bereits etablierten wissenschaftlichen Disziplin und damit den Fortbestand ihrer jeweiligen Erkenntnishaltung. Ausschlaggebend für das Selbstverständnis als Disziplin ist aber weder der bearbeitete Gegenstandsbereich (den die CP ja ohnehin weitgehend mit der traditionellen Literaturwissenschaft gemein hat), noch ihre Institutionalisierung. Es ist vielmehr die wissenschaftliche Erkenntnishaltung selbst, die dieses Selbstverständnis und damit auch die Wahrnehmung von außen bestimmt. Entscheidend ist insofern, mit welcher Erkenntnishaltung Computerphilologen in ihren zentralen Arbeitsbereichen an ihren Forschungsgegenstand herantreten. Zu diesen Arbeitsbereichen rechnen wir im wesentlichen die folgenden:

1. Archivierung, Edition, Dissemination von Texten im elektronischen Format;
2. quantitativ-lexematische bzw. qualitativ-semantische Analyse von Texten;
3. Annotation, Kommentierung, wissenschaftliche Interpretation und Analyse von Texten (Tagging, Kodierung, Modellierung);
4. Methodenreflexion und Theoriebildung.

#### 4. Ein Beispiel aus der CP-Praxis

Ich möchte jetzt ein Beispiel aus meiner eigenen computerphilologischen Forschungspraxis skizzieren, das unter den Arbeitsbereich drei fällt. In diesem Projekt wurde untersucht, wie und aufgrund welcher Textelemente und –Merkmale wir als Leser narrativer Texte eine Vorstellung von der Handlung gewinnen, die sich in einer erzählten Welt vollzieht.<sup>[31]</sup>

Der methodische Ansatz verknüpfte dabei Elemente aus der literaturwissenschaftlichen Erzähltheorie und Narratologie mit

computerphilologischer Textauszeichnung und einem Verfahren, das die Fähigkeit von Rechnern zur kombinatorischen Analyse von Textdaten nutzt. Der theoretische Fokus lag dabei auf dem Konzept der ‚Episode‘, also der kleinsten geschlossenen Einheit einer erzählten Handlung. Eine Episode kann man definieren als ein Interpretationskonstrukt des Lesers, der mittels dieses Konstrukts eine Reihe distinkter Sachverhalte in der fiktionalen Welt zu einer Serie miteinander verbundener Ereignisse verknüpft und damit dem Geschehen eine bestimmte Zielrichtung, wenn nicht gar einen Sinn beimißt. Nach dieser Auffassung *haben* Texte eigentlich keine Handlung, sonder stellen vielmehr das Geschehensmaterial zur Verfügung, um eine Handlung in ihnen *lesen* zu können. Dieser Prozeß hat den konkreten Text zur Basis; er bezieht jedoch darüber hinaus das Weltwissen des Lesers ein und verläuft in Abhängigkeit von bestimmten logischen und semiotischen Regeln, die den Status von *de facto*-Universalien des Lesens von Handlung besitzen. Soweit also die theoretischen Prämissen.

Neben der Formulierung neuer Definitionen für die Kernbegriffe Ereignis, Episode und Handlung konzentrierte sich das Projekt auf die Entwicklung von zwei Softwareanwendungen, mit denen das theoretische Modell in der Praxis der Literaturanalyse zur Anwendung gebracht werden konnte. Das erste Tool, EventParser genannt, ist ein Mark-up Werkzeug, das die Identifizierung und standardkonforme Auszeichnung von einzelnen Ereignissen im narrativen Text erleichtert. Die so erzeugten Textdaten wurden dann mit einem zweiten eigens entwickelten Programm ausgewertet, das den Namen EpiTest trägt. EpiTest verwendet einen kombinatorischen Algorithmus, mit dem alle theoretisch möglichen Episoden generiert werden, die man aus den von Lesern identifizierten Ereignissen bilden kann. Die Menge dieser per Computer generierten *virtuellen* Episoden übersteigt um ein erhebliches die Anzahl der *manifesten* Episoden, die wir als Leser normalerweise wahrnehmen. Man kann die Zahl der virtuellen Episoden und ihre Relation zur Zahl der manifesten Episoden damit als einen Indikator für das *Handlungspotential* nutzen, das einen spezifischen Text im Vergleich zu anderen Texten auszeichnet.

Der Text, an dem Theorie und Software schließlich erprobt wurden, waren Goethes *Unterhaltungen deutscher Ausgewanderten* (1795), ein aus sechs Einzelerzählungen und von einer Rahmenhandlung umschlossener Novellenroman, der die Literaturwissenschaften insbesondere deshalb beschäftigt hat, weil die in diesen Einzelerzählungen enthaltenen Handlungen von der trivialsten Gespenstergeschichte bis zu einem den Leser verwirrenden symbolischen *Märchen* reichen, dessen Handlung man, salopp gesagt, nur noch schwerlich ‚auf die Reihe‘ bringen kann. Die Untersuchung des Gesamttextes mit EventParser und EpiTest konnte zeigen, *warum* sich dieser so uneinheitliche Lektüreeindruck ergibt: es hat etwas mit der semiotischen Anschlußfähigkeit der von uns gelesenen individuellen Ereignisse zu tun, wie komplex und umfassend die von uns anschließend generierten komplexen Handlungskonstrukte jeweils ausfallen. Dieser Befund läßt sich in der Form eines Graphen darstellen, der das integrative Potential der Einzelerzählungen an drei Parametern (wieviele Ereignisse lassen sich prozentual zu Episoden

verknüpfen; wieviele Episoden kann man ihrerseits zu einer durchgehenden Handlung integrieren; wie hoch ist schließlich das Handlungspotential?) mißt:



Text segment	(%) EVENT integration	(%) EPISODE integration	ACTION product
<i>Antonelli</i>	26.95	27.40	738.48
<i>Bassompierre</i>	84.25	39.19	3301.71
<i>Klopffgeist</i>	16.27	35.29	574.16
<i>Prokurator</i>	59.06	14.65	865.20
<i>Ferdinand</i>	55.81	18.71	1044.04
<i>Märchen</i>	9.70	1.49	14.44
<b>Average</b>	<b>42.01</b>	<b>22.79</b>	<b>1089.67</b>
<i>Unterhaltungen total</i>	6.34	6.50	41.22

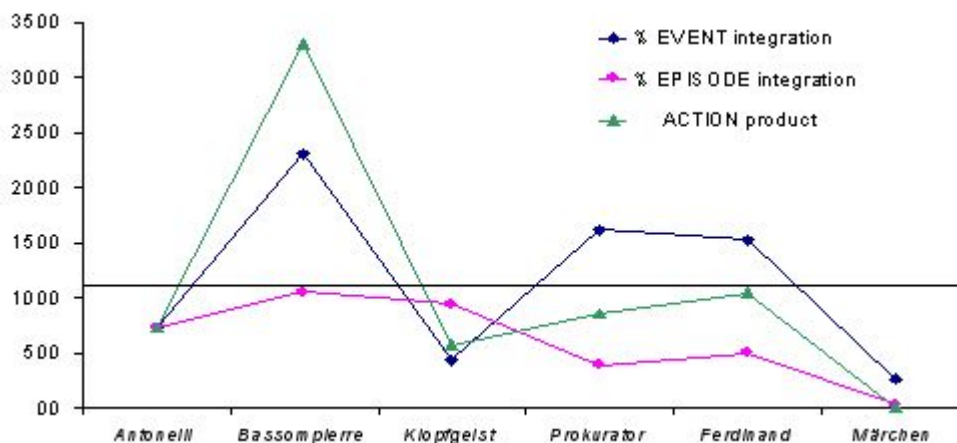


Abbildung 5: Kombinatorische Analyse des Handlungspotentials mit EpiTest nach Meister 2003.

Die so generierten Verlaufskurven wurden abschließend in Beziehung gesetzt zu den verschiedenen Textdeutungen, die in der Rezeptions- und Forschungsgeschichte zu Goethes *Unterhaltungen* dokumentiert sind. Das resultierte dann in der Formulierung einer neuen Textdeutung, die den epistemologischen Gehalt des Goethetextes in der Vordergrund stellte.

## 5. Zur philologischen Relevanz der CP

Methodisch gesehen ist es ganz wesentlich dieser letzte Schritt, der eine computerphilologische Studie überhaupt erst als einen genuine Beitrag zur Philologie ausweist: die computerphilologischen Ergebnisse müssen nämlich als philologisch relevante Befunde interpretierbar sein. Mit anderen Worten, sie müssen entweder neue Grundlagen im Gegenstandsfeld schaffen (Textedition), oder einen methodisch gesicherten Ansatzpunkt für die Formulierung neuer Fragen und Deutungshypothesen (Heuristik) bereitstellen, oder aber uns in die Lage versetzen, zur Evaluation bestehender wie zur eventuellen Formulierung neuer Deutungen eines Textgegenstandes bzw. Textkorpus zu gelangen (Kritik und Hermeneutik).

Mit dieser schließlichen Hin- und Rückwendung zu dem Problemtypus, der die traditionelle Philologie bestimmt, wird zunächst

die Anbindung der CP an den methodischen Kontext der im wesentlichen hermeneutisch orientierten Textwissenschaften gesichert. Zugleich jedoch wird auch ein umfassenderer Beitrag zur Wissenschaftsentwicklung geleistet, nämlich eine konkrete Vermittlung zwischen den beiden Wissenskulturen der Literatur(wissenschaften) einerseits, der Naturwissenschaften andererseits, die sich nach Charles Percy Snow durch je eigene wissenschaftliche Epistemologien und Methodologien auszeichnen.[32] Denn die CP hat offenkundig Anteil an beiden Methodologien: was ihre konkreten Arbeitsformen angeht, so ist sie stärker als die klassischen Textwissenschaften am Paradigma der empirischen Wissenschaften orientiert. In dieser Hinsicht sind ihre Hauptmerkmale:

- die Be- / Verarbeitung von objektiv quantifizierbaren Textdaten;
- deren algorithmengesteuerte Bearbeitung und Analyse;
- die non-ambige, immanent bedeutungsneutrale Transformation von Input in Output;
- die Evaluation ihrer Ergebnisse nach Maßgabe eines statistischen bzw. differentiellen Relevanzkriteriums.

Andererseits aber bleibt die CP dem hermeneutischen Grundinteresse der Literaturwissenschaften verpflichtet, denn die Relevanz ihrer Ergebnisse bemißt sich in letzter Instanz an der Fähigkeit zur

- Erarbeitung von subjektiv qualifizierbaren Textinterpretationen;
- Exemplifikation neuer, origineller Interpretationsansätze;
- selbstkritischen Reflektion ihrer Ergebnisse nach Maßgabe übergeordneter ästhetischer bzw. lebensweltlicher Fragestellungen.

Das Projekt einer ambitionierten, sich selbst nicht nur instrumentell verstehenden CP, die eine eigene methodische Identität entwickelt, ist derzeit noch eher Vision als Realität. Praktisch realisieren lassen wird es sich nur in Form von Forschungsvorhaben, die ihren methodisch-systematischen Kontext umfassender reflektieren. Die nachfolgende Abbildung skizziert eine systematisch-genetische Matrix der Komponenten und Prozesse, die für solcherart Forschungsprojekte kennzeichnend ist. Ein idealtypisches Projekt würde dabei alle in der rechten Spalte aufgelisteten Phasen von unten nach oben durchlaufen:

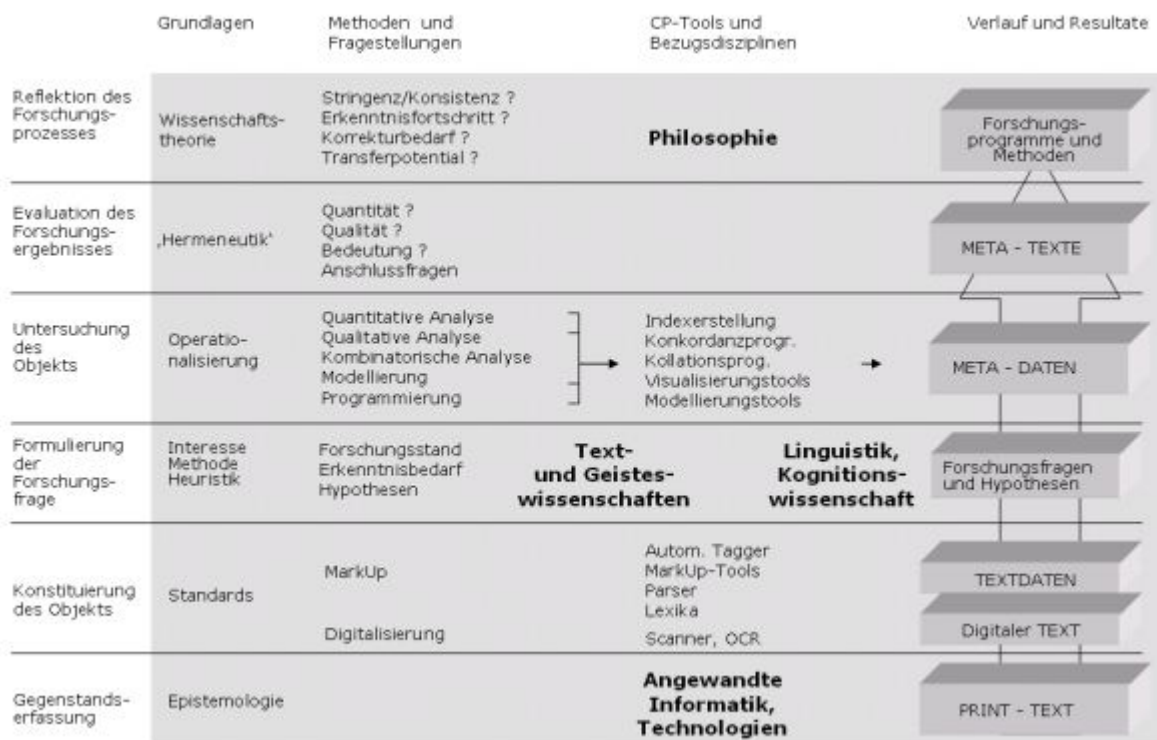


Abbildung 6: Systematisch-genetische Matrix für CP-Projekte

CP in diesem umfassenderen Sinne wäre damit denkbar weit entfernt von dem Projekt der ‚Literary Engine‘, mit dem sich Lemuel Gulliver vor mehr als 275 Jahren konfrontiert sah, und sie wäre auch wesentlich mehr als ein bloßes Hilfsmittel, mit dessen Hilfe sich traditionelle Philologie quantitativ optimieren ließe. Eine ambitionierte CP zielte vielmehr auf die Überwindung einer grundlegenden methodischen

Beschränkung der Philologie, die Heinrich von Kleist wie folgt charakterisiert hat:

Man könnte die Menschen in zwei Klassen abteilen: in solche, die sich auf eine Metapher, und in solche (2), die sich auf eine Formel verstehen. Deren, die sich auf beides verstehen, sind zu wenige; sie machen keine Klasse aus.

## Weiterführende Literatur

An Stelle einer – in diesem Forschungsbereich naturgemäß schnell überholten – Auswahlbibliographie sei hingewiesen auf vier im Internet zugängliche und laufend aktualisierte Verzeichnisse:

## Deutsch

Forum Computerphilologie (Universität München) –  
<http://computerphilologie.uni-muenchen.de/>  
 Arbeitsstelle Computerphilologie (Universität Hamburg) –  
<http://www.c-phil.uni-hamburg.de/>

## Englisch

Centre for Computing in the Humanities (King's College London) -  
<http://www.kcl.ac.uk/humanities/cch/>  
 Center for Electronic Texts in the Humanities (Rutgers University) -  
<http://www.ceth.rutgers.edu/index.html>

- 
- [1] Elektronischer Pre-Print, Version 03.09.2003. Der Beitrag ist zur Publikation vorgesehen in: Segeberg, Harro; Winko, Simone (Hg.): *Literarität und Digitalität. Zur Zukunft der Literatur*. (= Texte einer Hamburger Ringvorlesung im Wintersemester 2002/2003.) München (voraussichtl. Erscheinungsdatum: 2004)
- [2] Zur Würdigung von Lullus als Vorläufer der Kognitiven Semantik siehe Wolfgang Wildgen, 'From Lullus to Cognitive Semantics: The Evolution of a Theory of Semantic Fields'. Vortrag auf dem *Twentieth World Congress of Philosophy* in Boston, Massachusetts 10-15. August 1998: <http://www.bu.edu/wcp/Papers/Cogn/CognWild.htm> (19.08.2003). - Dem Lullischen und Leibnizschen kombinatorischen Ansatz verpflichtet ist u.a. die sog. 'Literary Machine'. Dabei handelt es sich um eine kommerzielle Software, die sich als 'dynamic archive and an idea management tool aimed at creative thinking' offeriert: [http://www.sommestad.com/LM\\_1\\_1.htm](http://www.sommestad.com/LM_1_1.htm) (19.08.2003)
- [3] CP wird bislang nur in wenigen Ausnahmefällen an deutschen Universitäten gelehrt. Umso wichtiger war deshalb bei der Vorbereitung dieses Beitrags der Gedankenaustausch mit interessierten Kollegen. Dank sagen möchte ich an dieser Stelle den Mitgliedern der Hamburger *Arbeitsstelle für Computerphilologie*, darunter insbesondere Walther von Hahn und Rolf Krause, mit dem mich u.a. ein aktuelles computerphilologisches Forschungsprojekt verbindet. Ein ganz besonderer Dank aber gilt unseren Studierenden: Wer in einem interdisziplinären, sich noch formierenden Feld wie der CP mit Gewinn arbeiten will, muß bereit sein, disziplinäre wie methodische Scheuklappen abzulegen. Nichts hält dazu besser an als die erbarmungslose Neugier, die unerschütterliche Kritikbereitschaft und die lebensfrohe Skepsis, die in einem Seminar regieren, sobald motivierte Studentinnen und Studenten einem die Kurbel aus der Hand nehmen!
- [4] Die nachfolgende historischen Rekonstruktion orientiert sich an Michael Fraser, *A Hypertextual History of Humanities Computing* (1996). <http://users.ox.ac.uk/~ctitext2/history/> (25.08.2003).
- [5] Siehe Roberto Busa: 'Thomae Aquinatis Opera Omnia cum Hypertextibus in CD-ROM'; in: *Cristianesimo nella Storia* vol. 18, Bologna 1997: 397-404 .
- [6] Zit. nach Michael Fraser, *A Hypertextual History of Humanities Computing* (1996). <http://info.ox.ac.uk/ctitext/history/pioneer.html> (25.08.2003)
- [7] Zum CETEDOC siehe <http://www.rice.edu/fondren/erc/resources/cetedoc.html> (19.08.2003)
- [8] Zit. nach M.Fraser, a.a.O. (= Anm. 5)
- [9] Zur ALLC siehe <http://www.allc.org/> ; zur ACH siehe <http://www.ach.org/> (19.08.03)
- [10] Siehe <http://www.princeton.edu/~mccarty/humanist/> (20.08.2003)
- [11] Paul Tombeur: 'Research carried out at the Centre de Traitement Electronique des Documents of the Catholic University of Louvain'; in AJ Aitken, RW Bailey and N. Hamilton-Smith, eds. *The Computer and Literary Studies*. Edinburgh (Edinburgh University Press), 1974:340
- [12] Zitiert nach Michael Fraser, *A Hypertextual History of Humanities Computing: Towards Independence from the Scientist*. <http://users.ox.ac.uk/~ctitext2/history/independ.html> (19.08.2003)
- [13] Zu TEI siehe <http://www.tei-c.org> (21.08.2003)
- [14] Mittlerweile hat die Entwicklung von TEI-Softwaretools die standardkonforme Textauszeichnung allerdings deutlich erleichtert und damit auch für non-kommerzielle Projekte bezahlbar gemacht.
- [15] Zum IATH siehe <http://www.iath.virginia.edu/> (21.08.2003)
- [16] Zum *Speculative Computing Laboratory* siehe <http://eotpac.clas.virginia.edu/speclab/index.html>; zum ETC siehe dagegen <http://etext.lib.virginia.edu/> (21.08.2003)
- [17] Siehe die Konferenz *The Humanities Computing Curriculum / The Computing Curriculum in the Arts and Humanities*, November 9-10, 2001, Nanaimo, British Columbia, Canada: <http://web.mala.bc.ca/siemensr/HCCcurriculum/> (21.08.2003)
- [18] Zu TAPOR siehe <http://huco.ualberta.ca/Tapor/> (21.08.2003)
- [19] Siehe <http://www.c-phil.uni-hamburg.de> (25.08.2003)
- [20] Zum Stand der HC-Entwicklung in Europa siehe die ACO\*HUM-Studie von 1999: <http://www.hit.uib.no/AcoHum/aco-hum.html> (21.08.2003).
- [21] <http://www.kcl.ac.uk/humanities/cch/alk/imhc/> (21.08.2003)
- [22] Siehe <http://www.hki.uni-koeln.de/> (21.08.2003)
- [23] Siehe <http://www.phf.uni-rostock.de/imd/index.htm> (21.08.2003)
- [24] Siehe <http://www.geschichte.hu-berlin.de/bereiche/histinf/> (21.08.2003)
- [25] Zu TUSTEP siehe <http://www.uni-tuebingen.de/zdv/tustep/>. Mit der Emeritierung Otts im Jahr 2002 scheint allerdings das Schicksal der Tübinger Arbeitsstelle besiegelt worden zu sein. Siehe hierzu <http://www.uni-tuebingen.de/uni/qvr/03fl1/03fl1-06-01.html> (21.08.2003)
- [26] Fotis Jannidis: 'Was ist Computerphilologie?' In: *Jahrbuch für Computerphilologie*, Bd.1, 1990:40f
- [27] Nach einer Powerpointfolie aus den Materialien zur Vorlesung Walter von Hahn, 'Einführung in die Computerphilologie', Universität Hamburg, Wintersemester 2001/2002.

- [28] *Jahrbuch für Computerphilologie*, Bd.4, 2002. Das Jahrbuch ist komplett online zugänglich unter <http://www.computerphilologie.uni-muenchen.de/jahrbuch/jb4-content.html> (25.08.2003)
- [29] Guntau, Martin / Laitko, Hubert: "Entstehung und Wesen wissenschaftlicher Disziplinen". In: dies.(Hgg.): *Der Ursprung der modernen Wissenschaften. Studien zur Entstehung wissenschaftlicher Disziplinen*. Berlin (Akademie Verlag) 1987, 17-89.
- [30] *ibid.*:35
- [31] Siehe Jan Christoph Meister, *Computing Action. A Narratological Approach. With a Foreword by Marie-Laure Ryan*. Berlin, New York (de Gruyter) 2003. On-line Informationen zu dem Projekt siehe unter <http://www.narratology.net/html/de010g.html> (21.08.2003)
- [32] Charles Percy Snow, *Literarische und naturwissenschaftliche Intelligenz*. Stuttgart 1967.